

Protein databases

Henrik Nielsen
(Thomas Nordahl)

Protein databases, historical background

Swiss-Prot, <http://www.expasy.org/sprot/>

Established in 1986 in Switzerland

ExPASy (Expert Protein Analysis System)

Swiss Institute of Bioinformatics (SIB) and European Bioinformatics Institute (EBI)

PIR, <http://pir.georgetown.edu/>

Established in 1984

National Biomedical Research Foundation, Georgetown University, USA

In 2002 merged into:

UniProt, <http://www.uniprot.org/>

A collaboration between SIB, EBI and Georgetown University.

VSMGLDAVDE SSMTGSFGGS NAQTSTEEVS QDSTDIMALL DNNMLGSMGD
TASITPE TKRNDN VEELEDELQI ANVPGAGPL PACFFAQML
KIHIEFVAVN NQVAVN PSLNVTGY SVEEQDIFL
NPPFATERYEN APRIFETVTVGHAHI AGSKFAPNPN
QS...SLW FSSSSASN PARSREDA ELRREEEA ENDEAQXQM
UniProt
the universal protein resource



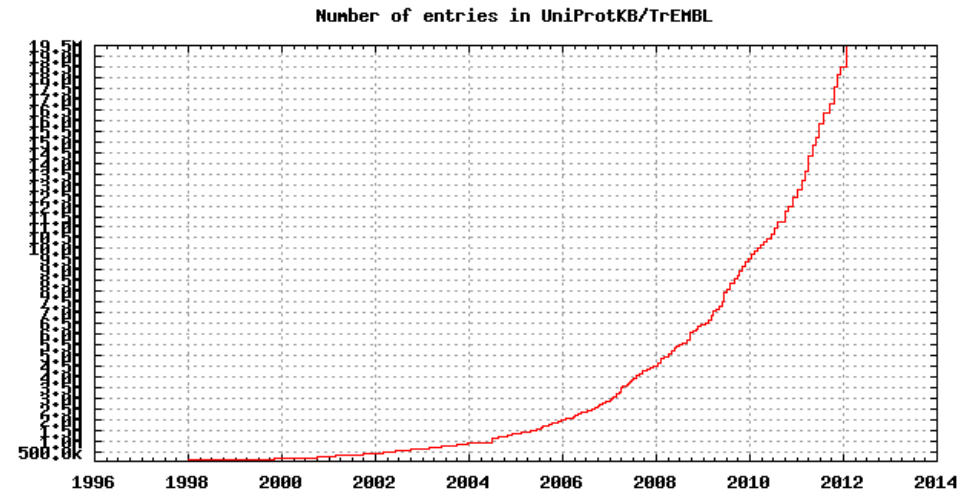
UniProt Knowledgebase Release 2012_01 (25-Jan-12)
consists of:

UniProtKB/Swiss-Prot: Annotated manually (*curated*)
534,242 entries

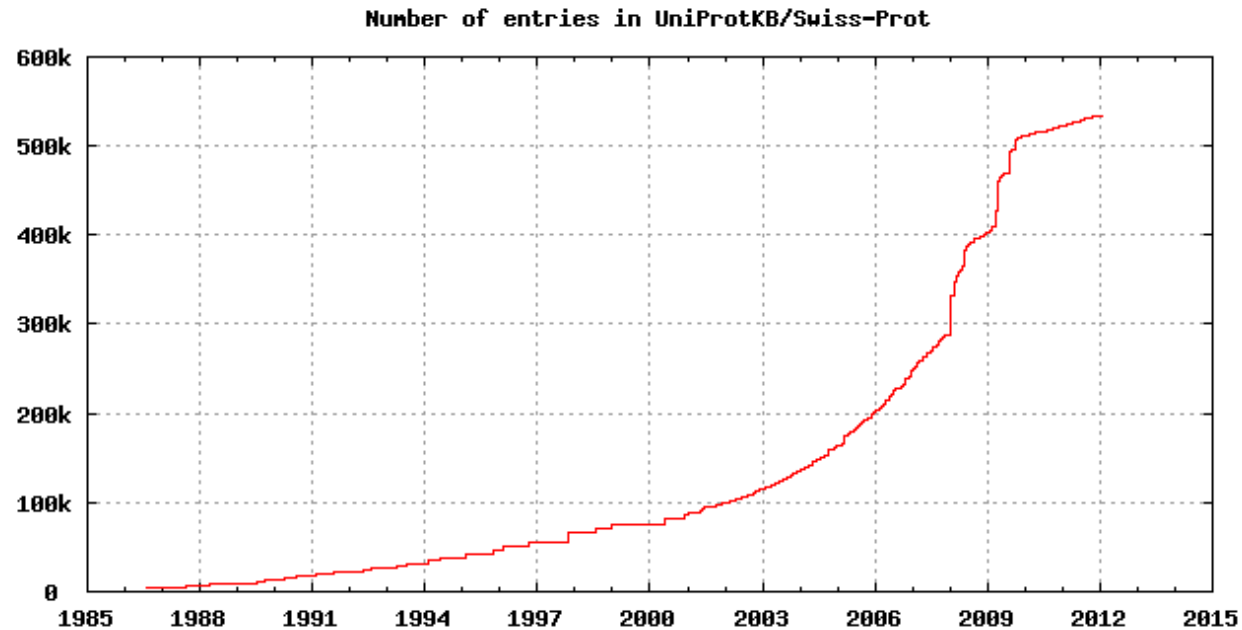
UniProtKB/TrEMBL: Computer annotated
19,434,245 entries

Growth of UniProt







































TrEMBL



Swiss-Prot



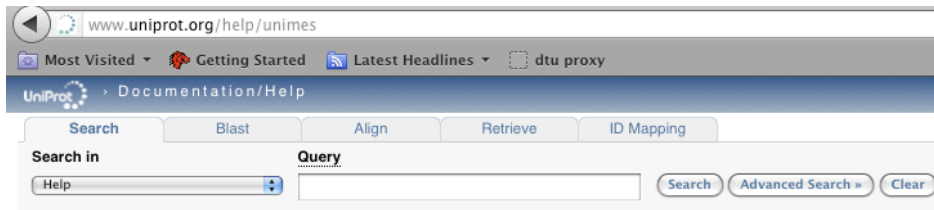
Camera db - <https://portal.camera.calit2.net/>

| Reads | Metadata | Assemblies | Peptides | Project Name |
|---|---|---|----------|--|
|  |  |  | | 13 Healthy Human Gut Metagenomes |
|  |  | | | A core gut microbiome in obese and lean twins |
|  |  | | | AAP gene operon study of SAR11 clades |
|  |  |  | | Acid Mine Drainage Metagenome |
|  |  | | | Alternative splicing in the choanoflagellate M. bre |
|  |  | | | Alvinella pompejana Epibiont Metagenome |
|  |  | | | Analysis of a viral metagenomic library from 200 m depth in Monterey Bay, California constructed by direct shotgun cloning |
|  |  | | | Antarctica Aquatic Microbial Metagenome |
|  |  | | | Bacterial Diversity in Soils under Artemisia sieberi and Noaea mucronata plants in the Negev Desert |
|  |  | | | Bermuda Oceanic Microbial Observatory Course |
|  |  | | | Bloomer DOM addition experiment |
|  |  | | | Bloomer DSW addition experiment |
|  |  | | | Botany Bay Metagenomes |
|  |  | | | Canine fecal metagenome |
|  |  | | | Human Distal Gut Biome project |
|  |  | | | Human Microbiome Project |
|  |  | | | Hydrothermal Vent Metagenome |
|  |  | | | Ice Metagenome of the Northern Schneeferner |

UniProt Knowledgebase Release 2012_01 (25-Jan-12)
consists of:

UniProtKB/Swiss-Prot: Annotated manually (*curated*)
534,242 entries

UniProtKB/TrEMBL: Computer annotated
19,434,245 entries



Unimes db – metagenomics data
<http://www.uniprot.org/help/unimes>

UniMES


UniMES

The UniProt Metagenomic and Environmental Sequences (UniMES) database is a repository specifically developed for metagenomic and environmental data. We provide UniMES clusters in order to obtain complete coverage of sequence space at different resolutions.

UniMES clusters

Clustered sets of sequences are available at two resolutions: 100% (unimes_cluster100.fasta) and >90% (unimes_cluster90.fasta). In unimes_cluster100.fasta, identical sequences and subfragments from unimes.fasta are placed into a single cluster. The unimes_cluster90.fasta is built by clustering unimes_cluster100.fasta representative sequences (the longest sequence in a cluster) using the CD-HIT algorithm (Li W., Jaroszewski L., and Godzik A., Bioinformatics, 17: 282-283, 2001) such that each cluster is composed of sequences that have at least 90% sequence identity, to the representative sequence. Only the representative sequences of the clusters are present in these files.

UniProt/TrEmbl


UniProtKB
Downloads · Contact · Documentation

Search
Blast
Align
Retrieve
ID Mapping *

Search in

Protein Knowledgebase (UniProtKB) ▾

Query

GDNF

Search

Advanced Search »

Clear

1 - 25 of 551 results for **GDNF** in UniProtKB sorted by **score** descending

Browse by [taxonomy](#), [keyword](#), [gene ontology](#), [enzyme class](#) or [pathway](#) | [Reduce sequence redundancy to 100%, 90% or 50%](#)

Page 1

Results Customize

- Show only [reviewed \(58\)](#) ★ (UniProtKB/Swiss-Prot) or [unreviewed \(493\)](#) ★ (UniProtKB/TrEMBL) entries
- Restrict term "gdnf" to [protein family \(13\)](#), [gene name \(58\)](#), [protein name \(95\)](#), [web resource \(1\)](#)

| Entry | Entry name | Status | Protein names | Gene names | Organism |
|---|--------------|--------|--|---------------------------------|--|
| <input type="checkbox"/> P39905 | GDNF_HUMAN | ★ | Glial cell line-derived neurotrophic factor | GDNF | Homo sapiens (Human) |
| <input type="checkbox"/> Q07731 | GDNF_RAT | ★ | Glial cell line-derived neurotrophic factor | Gdnf | Rattus norvegicus (Rat) |
| <input type="checkbox"/> P48540 | GDNF_MOUSE | ★ | Glial cell line-derived neurotrophic factor | Gdnf | Mus musculus (Mouse) |
| <input type="checkbox"/> Q06PM8 | GDNF_XENLA | ★ | Glial cell line-derived neurotrophic factor | gdnf | Xenopus laevis (African clawed frog) |
| <input type="checkbox"/> Q98TU0 | GDNF_DANRE | ★ | Glial cell line-derived neurotrophic factor | gdnf | Danio rerio (Zebrafish) (Brachydanio rerio) |
| <input type="checkbox"/> B7ZQI0 | B7ZQI0_XENLA | ★ | Gdnf protein | gdnf | Xenopus laevis (African clawed frog) |
| <input type="checkbox"/> O60609 | GFRA3_HUMAN | ★ | GDNF family receptor alpha-3 | GFRA3 UNQ339/PRO538/PRO3664 | Homo sapiens (Human) |
| <input type="checkbox"/> O35118 | GFRA3_MOUSE | ★ | GDNF family receptor alpha-3 | Gfra3 | Mus musculus (Mouse) |
| <input type="checkbox"/> Q8QGE9 | Q8QGE9_NIPNI | ★ | Glial cell line-derived neurotrophic factor G... | GDNF | Nipponia nippon (Crested ibis) (Ibis nippon) |
| <input type="checkbox"/> Q6UXV0 | GFRAL_HUMAN | ★ | GDNF family receptor alpha-like | GFRAL C6orf144 UNQ9356/PRO34128 | Homo sapiens (Human) |
| <input type="checkbox"/> P56159 | GFRA1_HUMAN | ★ | GDNF family receptor alpha-1 | GFRA1 GDNFRA RETL1 TRNR1 | Homo sapiens (Human) |

Content of UniProt Knowledgebase

- Amino acid sequences
 - Functional and structural annotations
 - Function / activity
 - Secondary structure
 - Subcellular location
 - Mutations, phenotypes (nsSNPs => cancer?)
 - Post-translational modifications
 - Origin
 - organism: Species, subspecies; classification
 - tissue
 - References
 - Cross references
-

Amino acid sequences

From where do you get amino acid sequences?

- Translation of nucleotide sequences (GenBank/EMBL/DDBJ)
 - Direct amino acid sequencing: *Edman degradation*
 - Mass spectrometry
 - 3D-structures
-



Search

Blast

Align

Retrieve

ID Mapping

Search in

Query

Protein Knowledgebase (UniProtKB) ↕

Human

Search

Advanced Search »

Clear

THE FIRST 10 YEARS OF UNIPROT

Saturday 8 September 2012, Basel, Switzerland

Celebrate UniProt's 10th anniversary with us! Renowned speakers from the fields of interactions & protein modelling, proteomics, protein structure & function, and genome analysis & annotation will highlight how protein databases are underpinning life sciences.

Submit abstracts for talks and posters to abstractsUP12@isb-sib.ch

WELCOME

The mission of **UniProt** is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

UniProtKB

Protein knowledgebase, consists of two sections:

- ★ Swiss-Prot, which is manually annotated and reviewed.

NEWS



UniProt release 2012_07 - Jul 11, 2012

To pee or not to pee | Removal of the cross-reference to CMR

- Statistics for UniProtKB:
[Swiss-Prot](#) · [TrEMBL](#)
- [Forthcoming changes](#)
- [News archives](#)

Follow @uniprot 380 followers

SITE TOUR



Learn how to make best use of the tools and data on this site.

Search

Blast

Align

Retrieve

ID Mapping *

Search in

Query

Protein Knowledgebase (UniProtKB) ▾

Human

Search

[Advanced Search »](#)

[Clear](#)

1 - 25 of 1,228,466 results for **Human** in UniProtKB sorted by **score** descending

[Browse by taxonomy, keyword, gene ontology, enzyme class or pathway](#) | [Reduce sequence redundancy to 100%, 90% or 50%](#)

[Download](#)

Page 1

of 49,139 | [Next »](#)

Results [Customize](#)

• Show only [reviewed \(45,542\)](#) ★ (UniProtKB/Swiss-Prot) or [unreviewed \(1,182,924\)](#) ★ (UniProtKB/TrEMBL) entries

• Restrict term "human" to [author \(3,916\)](#), [gene name \(1\)](#), [virus host \(116,614\)](#), [protein name \(2,587\)](#), [organism \(699,428\)](#), [strain \(4,125\)](#), [taxonomy \(701,839\)](#), [tissue \(375\)](#), [web resource \(595\)](#)

| Entry | Entry name | Status | Protein names | Gene names | Organism | Length |
|---|-------------|--------|--|----------------------------------|----------------------|--------|
| <input type="checkbox"/> Q9Y2H6 | FND3A_HUMAN | ★ | Fibronectin type-III domain-containing protel... | FND3A FNDC3 HUGO KIAA0970 | Homo sapiens (Human) | 1,198 |
| <input type="checkbox"/> Q8IZP9 | GPR64_HUMAN | ★ | G-protein coupled receptor 64 | GPR64 HE6 TM7LN2 | Homo sapiens (Human) | 1,017 |
| <input type="checkbox"/> Q6FHJ7 | SFRP4_HUMAN | ★ | Secreted frizzled-related protein 4 | SFRP4 FRPHE | Homo sapiens (Human) | 346 |
| <input type="checkbox"/> P42694 | HELZ_HUMAN | ★ | Probable helicase with zinc finger domain | HELZ DRHC KIAA0054 | Homo sapiens (Human) | 1,942 |
| <input type="checkbox"/> P08246 | ELNE_HUMAN | ★ | Neutrophil elastase | ELANE ELA2 | Homo sapiens (Human) | 267 |
| <input type="checkbox"/> Q95273 | CCDB1_HUMAN | ★ | Cyclin-D1-binding protein 1 | CCNDBP1 DIP1 GCIP HHM | Homo sapiens (Human) | 360 |
| <input type="checkbox"/> Q15334 | L2GL1_HUMAN | ★ | Lethal(2) giant larvae protein homolog 1 | LLGL1 DLG4 HUGL HUGL1 | Homo sapiens (Human) | 1,064 |
| <input type="checkbox"/> P61916 | NPC2_HUMAN | ★ | Epididymal secretory protein E1 | NPC2 HE1 | Homo sapiens (Human) | 151 |
| <input type="checkbox"/> Q6P2E9 | EDC4_HUMAN | ★ | Enhancer of mRNA-decapping protein 4 | EDC4 HEDLS | Homo sapiens (Human) | 1,401 |
| <input type="checkbox"/> Q9BWU0 | NADAP_HUMAN | ★ | Kanadaptn | SLC4A1AP HLC3 | Homo sapiens (Human) | 796 |

UniProt entry, formatted view

UniProt > UniProtKB Downloads · Contact · Documentation/Help

Search Blast * Align * Retrieve ID Mapping *

Search in Protein Knowledgebase (UniProtKB) Query Search Advanced Search > Clear

P01009 A1AT HUMAN ★ Reviewed, UniProtKB/Swiss-Prot
Last modified January 25, 2012. Version 180. History...

Contribute
Send feedback
Read comments (0) or add your own

Clusters with 100%, 90%, 50% identity Documents (6) Third-party data text xml rdf/xml gff fasta

Names · Attributes · General annotation · Ontologies · Interactions · Alt products · Sequence annotation · Sequences · References · Web links
Cross-refs · Entry info · Documents Customize order

Names and origin

| | |
|----------------------|--|
| Protein names | <p><i>Recommended name:</i> Alpha-1-antitrypsin</p> <p><i>Alternative name(s):</i> Alpha-1 protease inhibitor Alpha-1-antiproteinase Serpine A1</p> <p><u>Cleaved into the following chain:</u> 1. Short peptide from AAT Short name=SPAAT</p> |
| Gene names | <p>Name: SERPINA1 Synonyms: AAT, PI ORF Names: PRO0684, PRO2209</p> |
| Organism | Homo sapiens (Human) |
| Taxonomic identifier | 9606 [NCBI] |
| Taxonomic lineage | Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo |

Protein attributes

Accession #

UniProt ID

UniProt entry, text view (flat file)

```
ID      A1AT_HUMAN                      Reviewed:          418 AA.
AC      P01009; A6PX14; B2RDQ8; QOPVP5; Q13672; Q53XB8; Q5UOM1; Q7M4R2;
AC      Q86U18; Q86U19; Q96BF9; Q96ES1; Q9P1P0; Q9UCE6; Q9UCM3;
DT      21-JUL-1986, integrated into UniProtKB/Swiss-Prot.
DT      01-OCT-1996, sequence version 3.
DT      25-JAN-2012, entry version 180.
DE      RecName: Full=Alpha-1-antitrypsin;
DE      AltName: Full=Alpha-1 protease inhibitor;
DE      AltName: Full=Alpha-1-antiprotease;
DE      AltName: Full=Serpine A1;
DE      Contains:
DE          RecName: Full=Short peptide from AAT;
DE              Short=SPAAT;
DE      Flags: Precursor;
GN      Name=SERPINA1; Synonyms=AAT, PI; ORFNames=PRO0684, PRO2209;
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
OC      Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 1).
RX      MEDLINE=84107980; PubMed=6319097;
RA      Bollen A., Herzog A., Cravador A., Herion P., Chuchana P.,
RA      van der Straten A., Loriau R., Jacobs P., van Elsen A.;
RT      "Cloning and expression in Escherichia coli of full-length
RT      complementary DNA coding for human alpha 1-antitrypsin.";
RL      DNA 2:255-264(1983).
...
```

UniProt entry, formatted view

UniProt

> UniProtKB

Downloads · Contact · Documentation/Help

Search

Blast *

Align *

Retrieve

ID Mapping *

Search in

Protein Knowledgebase (UniProtKB)

Query

Search

Advanced Search >

Clear

P01009 (A1AT_HUMAN) ★ Reviewed, UniProtKB/Swiss-Prot

Last modified January 25, 2012. Version 180. [History...](#)

Contribute

[Send feedback](#)

[Read comments \(0\) or add your own](#)

Clusters with 100%, 90%, 50% identity | [Documents \(6\)](#) | [Third-party data](#)

[text](#) [xml](#) [rdf/xml](#) [gff](#) [fasta](#)

[Names](#) · [Attributes](#) · [General annotation](#) · [Ontologies](#) · [Interactions](#) · [Alt products](#) · [Sequence annotation](#) · [Sequences](#) · [References](#) · [Web links](#)

[Cross-refs](#) · [Entry info](#) · [Documents](#) · [Customize order](#)

Names and origin

| | |
|----------------------|---|
| Protein names | <div><i>Recommended name:</i> Alpha-1-antitrypsin</div> <div><i>Alternative name(s):</i> Alpha-1 protease inhibitor Alpha-1-antiproteinase Serpine A1</div> <div><i>Cleaved into the following chain:</i> 1. Short peptide from AAT Short name=SPAAT</div> |
| Gene names | <div>Name: SERPINA1</div> <div>Synonyms: AAT, PI</div> <div>ORF Names: PRO0684, PRO2209</div> |
| Organism | Homo sapiens (Human) |
| Taxonomic identifier | 9606 [NCBI] |
| Taxonomic lineage | Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo |

Protein attributes

UniProt entry, protein sequence

```
>sp|P01009|A1AT_HUMAN Alpha-1-antitrypsin OS=Homo sapiens GN=SERPINA1 PE=1 SV=3
MPSSVSWGILLLAGLCCLVPVSLAEDPQGDAQKTDTSHHQDHPTFNKITPNLAFAFS
LYRQLAHQSNSTNIFFSPVSIATAFAMLSLGTKADTHDEILEGLNFNLTETPEAQIHEGF
QELLRTLNQPDSQLQLTTGNGLFLSEGLKLVDKFLEDVKKLYHSEFTVNFGDTEEAKKQ
INDYVEKGTQGKIVDLVKELDRDTVFALVNYIFFKGKWERPFVVKDTEEDFHVDQVTTV
KVPMMKRLGMFNIQHCKKLSSWVLLMKYLGNTAIFFLPDEGKLQHLLENELTHDIITKFL
ENEDRRSASLHLPKLSITGTYDLKSVLGQLGITKVFSGADLSGVTEEAPLKLSKAVHKA
VLTIDEKGTAAAGAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK
```

UniProt entry, formatted view

UniProt

> UniProtKB

Downloads · Contact · Documentation/Help

Search

Blast *

Align *

Retrieve

ID Mapping *

Search in

Protein Knowledgebase (UniProtKB)

Query

Search

Advanced Search >

Clear

P01009 (A1AT_HUMAN) ★ Reviewed, UniProtKB/Swiss-Prot

Last modified January 25, 2012. Version 180. [History...](#)

Contribute

[Send feedback](#)

[Read comments \(0\) or add your own](#)

Clusters with 100%, 90%, 50% identity | [Documents \(6\)](#) | [Third-party data](#)

[text](#) [xml](#) [rdf/xml](#) [gff](#) [fasta](#)

[Names](#) · [Attributes](#) · [General annotation](#) · [Ontologies](#) · [Interactions](#) · [Alt products](#) · [Sequence annotation](#) · [Sequences](#) · [References](#) · [Web links](#)

[Cross-refs](#) · [Entry info](#) · [Documents](#) · [Customize order](#)

Names and origin

| | |
|----------------------|---|
| Protein names | <div>Recommended name: Alpha-1-antitrypsin</div> <div>Alternative name(s): Alpha-1 protease inhibitor Alpha-1-antiproteinase Serpine A1</div> <div>Cleaved into the following chain: 1. Short peptide from AAT Short name=SPAAT</div> |
| Gene names | <div>Name: SERPINA1</div> <div>Synonyms: AAT, PI</div> <div>ORF Names: PRO0684, PRO2209</div> |
| Organism | Homo sapiens (Human) |
| Taxonomic identifier | 9606 [NCBI] |
| Taxonomic lineage | Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo |

Protein attributes

UniProt entry, UniRef

UniProt

UniRef

Downloads · Contact · Documentation/Help

Search

Blast

Align

Retrieve

ID Mapping

Search in

Sequence Clusters (UniRef)

Query

member:P01009 identity:0.9

Search

Advanced Search »

Clear

1 result for member:P01009 AND identity:0.9 in UniRef

Map to UniProtKB or UniParc

Download

Page 1 of 1

Results

Customize

| Cluster ID | Status | Cluster name | Size | Cluster member(s) | Organisms | Length | Identity |
|--|--------|------------------------------|------|---|--|--------|----------|
| <input type="checkbox"/> UniRef90_P01009 | ★ | Cluster: Alpha-1-antitrypsin | 15 | P01009 E9KL23 P01009-3 G3QXZ8 H2NM53 UPI000046973E Q5RCW5 F6SEN6 G1S644 +5 | Homo sapiens (Human) ? Gorilla gorilla gorilla (Lowland gorilla) Pongo abelii (Sumatran orangutan) Macaca mulatta (Rhesus macaque) Nomascus leucogenys (Northern white-cheeked gibbon) (Hylobates leucogenys) Macaca fascicularis (Crab-eating macaque) (Cynomolgus monkey) Pan troglodytes (Chimpanzee) Papio anubis (Olive baboon) Chlorocebus aethiops (Green monkey) (Cercopithecus aethiops) | 418 | 90% |

Page 1 of 1

General annotation (Comments)

General annotation (Comments)

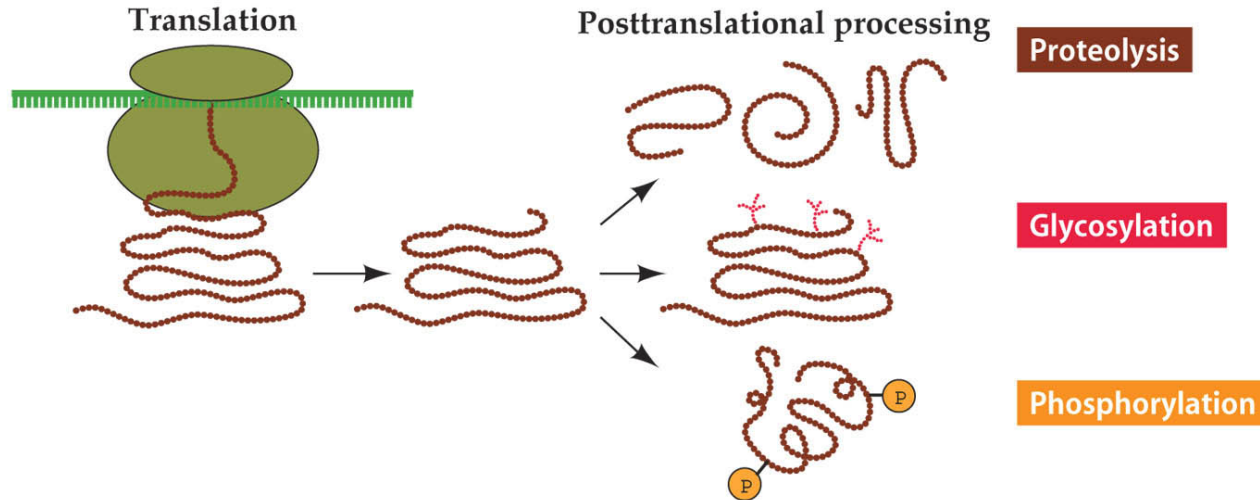
| | |
|---------------------------------|--|
| Function | <p>Inhibitor of serine proteases. Its primary target is elastase, but it also has a moderate affinity for plasmin and thrombin. Irreversibly inhibits trypsin, chymotrypsin and plasminogen activator. The aberrant form inhibits insulin-induced NO synthesis in platelets, decreases coagulation time and has proteolytic activity against insulin and plasmin. Ref.17 Ref.18 Ref.24</p> <p>Short peptide from AAT (SPAAT) is a reversible chymotrypsin inhibitor. It also inhibits elastase, but not trypsin. Its major physiological function is the protection of the lower respiratory tract against proteolytic destruction by human leukocyte elastase (HLE). Ref.17 Ref.18 Ref.24</p> |
| Subcellular location | <p>Secreted Ref.24.</p> <p><u>Short peptide from AAT</u>: Secreted > extracellular space > extracellular matrix Ref.24.</p> |
| Tissue specificity | Plasma. |
| Domain | The reactive center loop (RCL) extends out from the body of the protein and directs binding to the target protease. The protease cleaves the serpin at the reactive site within the RCL, establishing a covalent linkage between the carboxyl group of the serpin reactive site and the serine hydroxyl of the protease. The resulting inactive serpin-protease complex is highly stable. |
| Post-translational modification | <p>Several isomers are observed, resulting from the combination of different N-linked glycan structures and mature N-terminus. N-linked glycan at Asn-107 is alternatively di-antennary, tri-antennary or tetra-antennary, whereas glycan at Asn-70 is di-antennary with trace amounts of tri-antennary, and glycan at Asn-271 is exclusively di-antennary. The structure of the antennae is Neu5Ac(alpha1-6)Gal(beta1-4)GlcNAc attached to the core structure Man(alpha1-6)[Man(alpha1-3)]Man(beta1-4)GlcNAc(beta1-4)GlcNAc. Some antennae are fucosylated, which forms a Lewis-X determinant. Proteolytic processing may yield the truncated form that ranges from Asp-30 to Lys-418.</p> |
| Polymorphism | The sequence shown is that of the M1V allele which is the most common form of PI (44 to 49%). Other frequent alleles are: M1A 20 to 23%; M2 10 to 11%; M3 14 to 19%. |
| Involvement in disease | <p>Defects in SERPINA1 are the cause of alpha-1-antitrypsin deficiency (A1ATD) [MIM:613490]. A disorder whose most common manifestation is emphysema, which becomes evident by the third to fourth decade. A less common manifestation of the deficiency is liver disease, which occurs in children and adults, and may result in cirrhosis and liver failure. Environmental factors, particularly cigarette smoking, greatly increase the risk of emphysema at an earlier age. Ref.58 Ref.60 Ref.62</p> |
| Miscellaneous | The aberrant form is found in the plasma of chronic smokers, and persists after smoking is ceased. It can still be found ten years after smoking has ceased. |
| Sequence similarities | Belongs to the serpin family . |
| Sequence caution | <p>The sequence CAD62334.1 differs from that shown. Reason: Erroneous initiation. Translation N-terminally shortened.</p> <p>The sequence CAD62585.1 differs from that shown. Reason: Erroneous initiation. Translation N-terminally shortened.</p> |

General annotation (Comments)

General annotation (Comments)

| | |
|---------------------------------|--|
| Function | <p>Inhibitor of serine proteases. Its primary target is elastase, but it also has a moderate affinity for plasmin and thrombin. Irreversibly inhibits trypsin, chymotrypsin and plasminogen activator. The aberrant form inhibits insulin-induced NO synthesis in platelets, decreases coagulation time and has proteolytic activity against insulin and plasmin. Ref.17 Ref.18 Ref.24</p> <p>Short peptide from AAT (SPAAT) is a reversible chymotrypsin inhibitor. It also inhibits elastase, but not trypsin. Its major physiological function is the protection of the lower respiratory tract against proteolytic destruction by human leukocyte elastase (HLE). Ref.17 Ref.18 Ref.24</p> |
| Subcellular location | <p>Secreted Ref.24.</p> <p>Short peptide from AAT: Secreted > extracellular space > extracellular matrix Ref.24.</p> |
| Tissue specificity | Plasma. |
| Domain | The reactive center loop (RCL) extends out from the body of the protein and directs binding to the target protease. The protease cleaves the serpin at the reactive site within the RCL, establishing a covalent linkage between the carboxyl group of the serpin reactive site and the serine hydroxyl of the protease. The resulting inactive serpin-protease complex is highly stable. |
| Post-translational modification | <p>Several isomers are observed, resulting from the combination of different N-linked glycan structures and mature N-terminus. N-linked glycan at Asn-107 is alternatively di-antennary, tri-antennary or tetra-antennary, whereas glycan at Asn-70 is di-antennary with trace amounts of tri-antennary, and glycan at Asn-271 is exclusively di-antennary. The structure of the antennae is Neu5Ac(alpha1-6)Gal(beta1-4)GlcNAc attached to the core structure Man(alpha1-6)[Man(alpha1-3)]Man(beta1-4)GlcNAc(beta1-4)GlcNAc. Some antennae are fucosylated, which forms a Lewis-X determinant. Proteolytic processing may yield the truncated form that ranges from Asp-30 to Lys-418.</p> |
| Polymorphism | The sequence shown is that of the M1V allele which is the most common form of PI (44 to 49%). Other frequent alleles are: M1A 20 to 23%; M2 10 to 11%; M3 14 to 19%. |
| Involvement in disease | Defects in SERPINA1 are the cause of alpha-1-antitrypsin deficiency (A1ATD) [MIM:613490]. A disorder whose most common manifestation is emphysema, which becomes evident by the third to fourth decade. A less common manifestation of the deficiency is liver disease, which occurs in children and adults, and may result in cirrhosis and liver failure. Environmental factors, particularly cigarette smoking, greatly increase the risk of emphysema at an earlier age. Ref.58 Ref.60 Ref.62 |
| Miscellaneous | The aberrant form is found in the plasma of chronic smokers, and persists after smoking is ceased. It can still be found ten years after smoking has ceased. |
| Sequence similarities | Belongs to the serpin family . |
| Sequence caution | <p>The sequence CAD62334.1 differs from that shown. Reason: Erroneous initiation. Translation N-terminally shortened.</p> <p>The sequence CAD62585.1 differs from that shown. Reason: Erroneous initiation. Translation N-terminally shortened.</p> |

Post-translational modifications



Mange proteiner skal *modificeres* efter at de er syntetiseret for at blive aktive

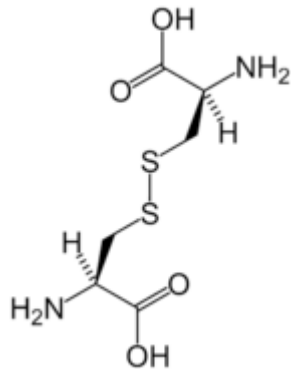
Proteolyse: Fraspaltning af *signalpeptider*, *propeptider* eller *initiator methionin*

Glykosylering: Gælder især proteiner på *overfladen* af celler. Spiller også en rolle i sortering af proteiner til *lysosomer*

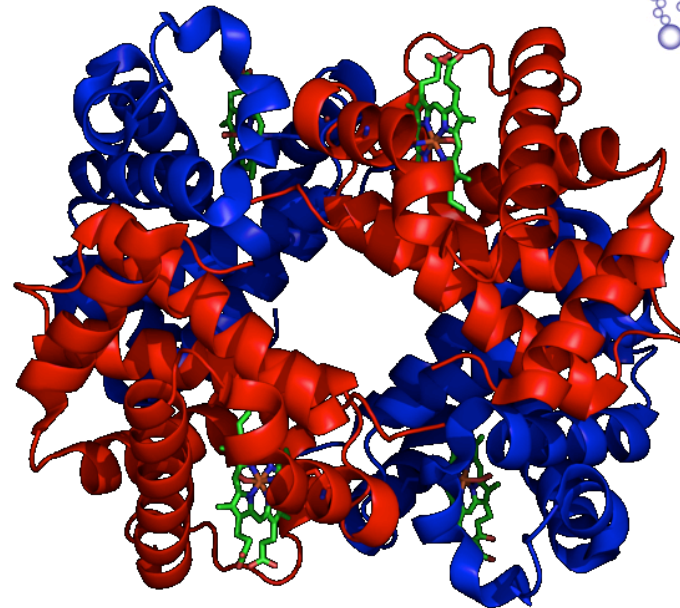
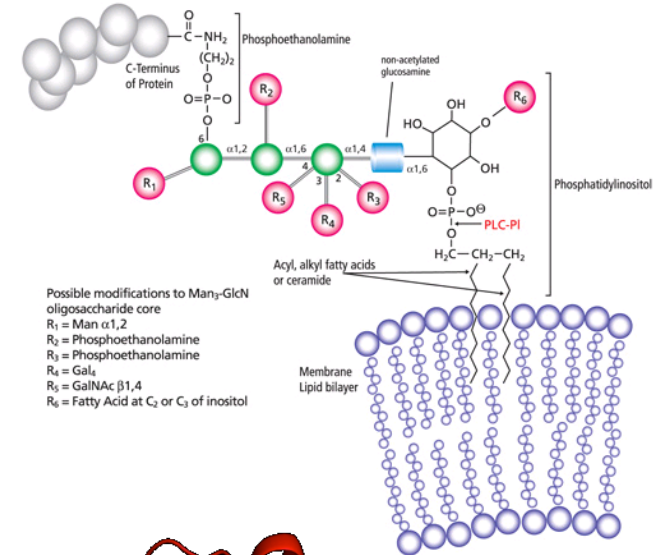
Phosphorylering: Ofte *reversibel*. Regulerer *aktiviteten* af mange enzymer

More post-translational modifications

- Lipid anchors
 - (e.g. GPI anchors)
- Disulfide bonds



- Prosthetic groups
 - (e.g. metal ions)



General annotation (Ontologies)

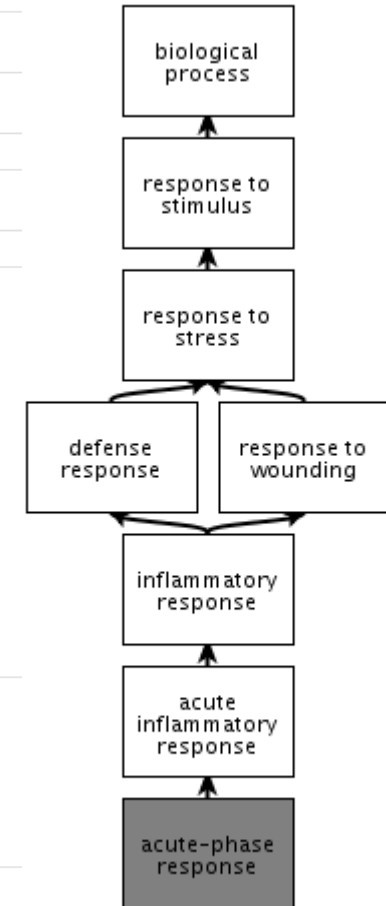
Ontologies

Keywords

| | |
|---------------------------|--|
| Biological process | Acute phase Blood coagulation |
| Cellular component | Extracellular matrix Secreted |
| Coding sequence diversity | Alternative splicing Polymorphism |
| Domain | Signal |
| Molecular function | Protease inhibitor Serine protease inhibitor |
| PTM | Glycoprotein |
| Technical term | 3D-structure Complete proteome Direct protein sequencing Reference proteome |

Gene Ontology (GO)

| | |
|--------------------|--|
| Biological process | acute-phase response Inferred from electronic annotation. Source: UniProtKB-KW platelet activation Traceable author statement. Source: Reactome platelet degranulation Traceable author statement. Source: Reactome regulation of proteolysis Inferred from Biological aspect of Ancestor. Source: RefGenome |
| Cellular component | extracellular space Inferred from mutant phenotype. Source: UniProtKB platelet alpha granule lumen Traceable author statement. Source: Reactome proteinaceous extracellular matrix Inferred from electronic annotation. Source: UniProtKB-SubCell |
| Molecular function | protease binding Inferred from physical interaction. Source: UniProtKB serine-type endopeptidase inhibitor activity Non-traceable author statement. Source: UniProtKB |



QuickGO - <http://www.ebi.ac.uk/QuickGO>

Sequence annotation (Feature Table)

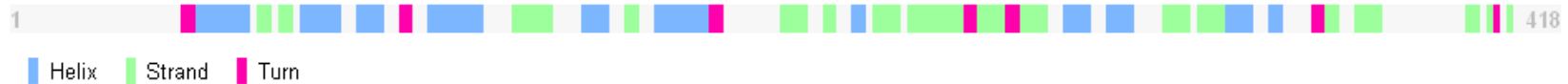
Sequence annotation (Features)

| | Feature key | Position(s) | Length | Description | Graphical view | Feature identifier |
|---------------------------------|----------------------|-------------|--------|---|----------------|--------------------|
| Molecule processing | | | | | | |
| <input type="checkbox"/> | Signal peptide | 1 – 24 | 24 | Ref.15 Ref.16 Ref.17 Ref.18 | | |
| <input type="checkbox"/> | Chain | 25 – 418 | 394 | Alpha-1-antitrypsin Ref.2 | | PRO_0000032377 |
| <input type="checkbox"/> | Peptide | 375 – 418 | 44 | Short peptide from AAT | | PRO_0000364030 |
| Regions | | | | | | |
| <input type="checkbox"/> | Region | 368 – 392 | 25 | RCL | | |
| Sites | | | | | | |
| <input type="checkbox"/> | Site | 382 – 383 | 2 | Reactive bond | | |
| Amino acid modifications | | | | | | |
| <input type="checkbox"/> | Modified residue | 256 | 1 | S-cysteinyln cysteine | | |
| <input type="checkbox"/> | Glycosylation | 70 | 1 | N-linked (GlcNAc...) (complex) Ref.19 Ref.26 Ref.27 Ref.28 Ref.29 Ref.30 Ref.31 Ref.32 | | |
| <input type="checkbox"/> | Glycosylation | 107 | 1 | N-linked (GlcNAc...) (complex) Ref.19 Ref.29 Ref.31 Ref.32 | | |
| <input type="checkbox"/> | Glycosylation | 271 | 1 | N-linked (GlcNAc...) (complex) Ref.19 Ref.26 Ref.27 Ref.28 Ref.29 Ref.31 Ref.32 | | |
| Natural variations | | | | | | |
| <input type="checkbox"/> | Alternative sequence | 307 – 418 | 112 | Missing in isoform 3. | | VSP_028890 |
| <input type="checkbox"/> | Alternative sequence | 356 – 418 | 63 | AVHKA...NPTQK → VRSP in isoform 2. | | VSP_028889 |
| <input type="checkbox"/> | Natural variant | 4 | 1 | S → L in Z-Wrexham. Ref.62 | | VAR_006978 |
| <input type="checkbox"/> | Natural variant | 26 | 1 | D → A in V-Munich. Ref.59 | | VAR_006979 |
| <input type="checkbox"/> | Natural variant | 37 | 1 | T → A. Corresponds to variant rs11558262 [dbSNP Ensembl]. | | VAR_051938 |
| <input type="checkbox"/> | Natural variant | 58 | 1 | A → T in M5-Karlsruhe. | | VAR_006980 |
| <input type="checkbox"/> | Natural variant | 63 | 1 | R → C in I. Ref.63 Corresponds to variant rs28931570 [dbSNP Ensembl]. | | VAR_006981 |

Isoforms:
Mostly due to
alternative splicing. May be
tissue specific.

Secondary structure (Feature Table)

Secondary structure



| | | | | |
|--------------------------------------|-----------|----|--|--|
| Details... | | | | |
| <input type="checkbox"/> Turn | 48 – 50 | 3 | | |
| <input type="checkbox"/> Helix | 51 – 68 | 18 | | |
| <input type="checkbox"/> Beta strand | 70 – 72 | 3 | | |
| <input type="checkbox"/> Beta strand | 74 – 76 | 3 | | |
| <input type="checkbox"/> Helix | 78 – 89 | 12 | | |
| <input type="checkbox"/> Helix | 94 – 103 | 10 | | |
| <input type="checkbox"/> Turn | 108 – 110 | 3 | | |
| <input type="checkbox"/> Helix | 113 – 127 | 15 | | |
| <input type="checkbox"/> Beta strand | 135 – 145 | 11 | | |
| <input type="checkbox"/> Helix | 152 – 160 | 9 | | |
| <input type="checkbox"/> Beta strand | 165 – 169 | 5 | | |
| <input type="checkbox"/> Helix | 174 – 188 | 15 | | |
| <input type="checkbox"/> Turn | 189 – 191 | 3 | | |
| <input type="checkbox"/> Beta strand | 206 – 215 | 10 | | |
| <input type="checkbox"/> Beta strand | 218 – 220 | 3 | | |
| <input type="checkbox"/> Helix | 224 – 226 | 3 | | |
| <input type="checkbox"/> Beta strand | 228 – 237 | 10 | | |
| <input type="checkbox"/> Beta strand | 239 – 256 | 18 | | |
| <input type="checkbox"/> Turn | 257 – 260 | 4 | | |
| <input type="checkbox"/> Beta strand | 261 – 268 | 8 | | |
| <input type="checkbox"/> Turn | 269 – 271 | 3 | | |
| <input type="checkbox"/> Beta strand | 272 – 279 | 8 | | |

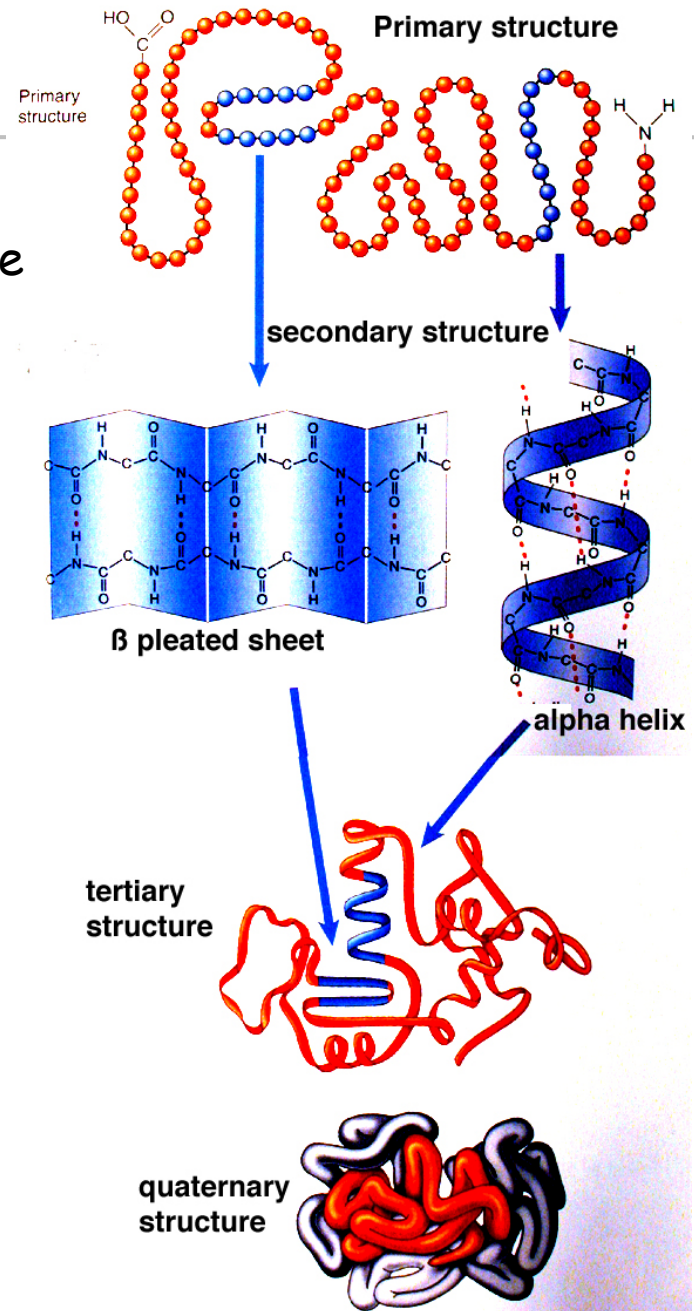
Protein structure

Primary structure: Amino acid sequence

Secondary structure:
"Backbone" hydrogen bonding
Alpha helix / Beta sheet / Turn

Tertiary structure: Fold, 3D coordinates

Quaternary structure: subunits



Evidence (Comments, Feature Table)







Q43495 (108_SOLLC) ★ Reviewed, UniProtKB/Swiss-Prot

Last modified March 2, 2010. Version 49.  [History...](#)

General annotation (Comments)

| | |
|-----------------------|---|
| Subcellular location | Secreted Potential |
| Tissue specificity | Stamen- and tapetum-specific. |
| Sequence similarities | Belongs to the A9/FIL1 family. |

Sequence annotation (Features)

| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier |
|---|--------------------------|--------|----------------------------|--|--------------------|
| Molecule processing | | | | | |
| <input type="checkbox"/> Signal peptide | 1 – 30 | 30 | Potential |  | |
| <input type="checkbox"/> Chain | 31 – 102 | 72 | Protein 108 |  | PRO_0000000238 |
| Amino acid modifications | | | | | |
| <input type="checkbox"/> Disulfide bond | 41 ↔ 77 | | By similarity |  | |
| <input type="checkbox"/> Disulfide bond | 51 ↔ 66 | | By similarity |  | |
| <input type="checkbox"/> Disulfide bond | 67 ↔ 92 | | By similarity |  | |
| <input type="checkbox"/> Disulfide bond | 79 ↔ 99 | | By similarity |  | |

Evidence/Confidence types

3 types of *non-experimental qualifiers* in Sequence annotation and General comment:

- *Potential*: Predicted using sequence analysis
 - *Probable*: Uncertain experimental evidence
 - *By similarity*: Predicted using sequence similarity
-

| Sequence | Length | Mass (Da) | Tools |
|---|--------------|-----------|--|
| <input type="checkbox"/> Isoform 1 [UniParc]. Last modified October 1, 1996. Version 3. Checksum: 7016555F273B7F16 | FASTA 418 | 46,737 | <input type="text" value="Blast"/> <input type="button" value="go"/> |

« Hide

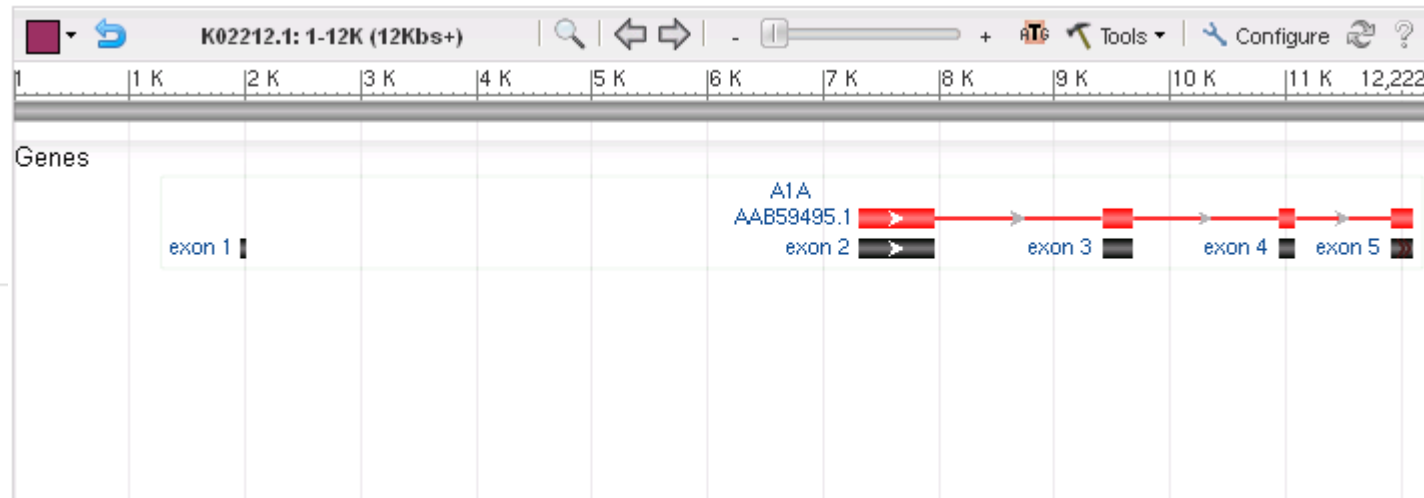
☐ **Isoform 3** [UniParc]. FASTA 306 34,755
 Checksum: 15C708E6C25CE0C4
[Show »](#)

Cross-references, nucleotide sequences

Sequence databases

- EMBL
- GenBank
- DDBJ

K01396 mRNA. Translation: AAB59375.1.
K02212 Genomic DNA. Translation: AAB59495.1.
X01683 mRNA. Translation: CAA25838.1.
M11465 mRNA. Translation: AAA51546.1.
J02619 Genomic DNA. Translation: AAA51547.1.
DQ682455 mRNA. Translation: ABG73380.1.
AM048838 Genomic DNA. Translation: CAJ15161.1.
AF113676 mRNA. Translation: AAF29581.1.
AF130068 mRNA. Translation: AAG35496.1.
BX161449 mRNA. Translation: CAD61914.1.
BX247968 mRNA. Translation: CAD62306.1.
BX248002 mRNA. Translation: CAD62334.1. Different initiation.
BX248257 mRNA. Translation: CAD62585.1. Different initiation.
AK315637 mRNA. Translation: BAG38005.1.
BT019455 mRNA. Translation: AAV38262.1.



Cross-references, 3D structure

3D structure databases

- PDBe
- RCSB PDB
- PDBj

| Entry | Method | Resolution (Å) | Chain | Positions | PDBsum |
|-------|--------|----------------|-------|-----------|--------|
| 1ATU | X-ray | 2.70 | A | 45-418 | [>] |
| 1D5S | X-ray | 3.00 | A | 44-377 | [>] |
| | | | B | 378-418 | [>] |
| 1EZU | X-ray | 2.60 | A | 48-382 | [>] |
| | | | B | 383-418 | [>] |
| 1HP7 | X-ray | 2.10 | A | 25-418 | [>] |
| 1IZ2 | X-ray | 2.20 | A | 25-418 | [>] |
| 1KCT | X-ray | 3.46 | A | 25-418 | [>] |
| 1OO8 | X-ray | 2.65 | A | 26-418 | [>] |
| 1OPH | X-ray | 2.30 | A | 26-418 | [>] |
| 1PSI | X-ray | 2.92 | A | 26-418 | [>] |
| 1QLP | X-ray | 2.00 | A | 26-418 | [>] |
| 1QMB | X-ray | 2.60 | A | 49-376 | [>] |
| | | | B | 377-418 | [>] |
| 2D26 | X-ray | 3.30 | A | 26-382 | [>] |
| | | | B | 383-418 | [>] |
| 2QUG | X-ray | 2.00 | A | 25-418 | [>] |
| 3CWL | X-ray | 2.44 | A | 25-418 | [>] |
| 3CWM | X-ray | 2.51 | A | 25-418 | [>] |
| 3DRM | X-ray | 2.20 | A | 26-418 | [>] |
| 3DRU | X-ray | 3.20 | A/B/C | 26-418 | [>] |
| 3NDD | X-ray | 1.50 | A | 46-372 | [>] |
| | | | B | 383-418 | [>] |
| 3NDF | X-ray | 2.70 | A | 46-381 | [>] |
| | | | B | 383-418 | [>] |
| 3T1P | X-ray | 3.90 | A | 48-418 | [>] |
| 7API | X-ray | 3.00 | A | 36-382 | [>] |
| | | | B | 383-418 | [>] |
| 8API | X-ray | 3.10 | A | 36-382 | [>] |
| | | | B | 383-418 | [>] |
| 9API | X-ray | 3.00 | A | 36-382 | [>] |
| | | | B | 383-418 | [>] |



Cross-references

Other databases linked from UniProt

(there are ~100 in total):

- Nucleotide sequences
 - 3D structure
 - Protein-protein interactions
 - Enzymatic activities and pathways
 - Gene expression (microarrays and 2D-PAGE)
 - Ontologies
 - Families and domains
 - Organism specific databases
-

Translation and Reading Frames

The genetic code

| | | Second letter | | | | |
|--------------|---|--|---|--|--|------------------|
| | | U | C | A | G | |
| First letter | U | <div>UUU</div> <div>UUC</div> Phenylalanine <div>UUA</div> <div>UUG</div> Leucine | <div>UCU</div> <div>UCC</div> <div>UCA</div> <div>UCG</div> Serine | <div>UAU</div> <div>UAC</div> Tyrosine <div>UAA</div> <div>UAG</div> Stop codon Stop codon | <div>UGU</div> <div>UGC</div> Cysteine <div>UGA</div> <div>UGG</div> Stop codon Tryptophan | U C A G |
| | C | <div>CUU</div> <div>CUC</div> <div>CUA</div> <div>CUG</div> Leucine | <div>CCU</div> <div>CCC</div> <div>CCA</div> <div>CCG</div> Proline | <div>CAU</div> <div>CAC</div> Histidine <div>CAA</div> <div>CAG</div> Glutamine | <div>CGU</div> <div>CGC</div> <div>CGA</div> <div>CGG</div> Arginine | U C A G |
| | A | <div>AUU</div> <div>AUC</div> <div>AUA</div> Isoleucine <div>AUG</div> Methionine; start codon | <div>ACU</div> <div>ACC</div> <div>ACA</div> <div>ACG</div> Threonine | <div>AAU</div> <div>AAC</div> Asparagine <div>AAA</div> <div>AAG</div> Lysine | <div>AGU</div> <div>AGC</div> Serine <div>AGA</div> <div>AGG</div> Arginine | U C A G |
| | G | <div>GUU</div> <div>GUC</div> <div>GUA</div> <div>GUG</div> Valine | <div>GCU</div> <div>GCC</div> <div>GCA</div> <div>GCG</div> Alanine | <div>GAU</div> <div>GAC</div> Aspartic acid <div>GAA</div> <div>GAG</div> Glutamic acid | <div>GGU</div> <div>GGC</div> <div>GGA</div> <div>GGG</div> Glycine | U C A G |

- Degenerate (*redundant*) but not ambiguous (4*4*4 combinations)
- *Almost* universal (deviations found in mitochondria)

Reading Frames 1

A piece of an mRNA-strand:

5' aug cccaagcugaauagcguagagggguuuucaucauuugaggacgauguaaa 3'

can be divided into triplets (*codons*) in three ways:

| | | | | | | | | | | | | | | | | | | | |
|---|--|-----|-----|--------------------------------------|--------------------------------------|-----|-----|--------------------------------------|-----|-----|-----|-----|--------------------------------------|-----|-----|-----|-----|-----|--|
| 1 | aug | ccc | aag | cug | aa | u | agc | gua | gag | ggg | uuu | uca | uca | uuu | gag | gac | gau | gua | u a a |
| | M | P | K | L | N | S | V | E | G | F | S | S | F | E | D | D | V | * | |
| 2 | ugc | cca | agc | uga | aua | g | cg | uag | agg | ggu | uuu | cau | cau | uug | agg | acg | aug | uau | |
| | C | P | S | * | I | A | * | R | G | F | H | H | L | R | T | M | Y | | |
| 3 | gcc | caa | gcu | gaa | uag | cgu | aga | ggg | guu | uuc | auc | auu | uga | gga | cga | ugu | aua | | |
| | A | Q | A | E | * | R | R | G | V | F | I | I | * | G | R | C | I | | |

Each possible set of triplets is called a *reading frame*.

Reading Frames 2

Since there are two strands in DNA, there are *six* possible reading frames in a piece of DNA (three in each direction):

| | | | | | | | | | | | | | | | | | | | |
|----|---|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----|----|
| 3 | A | Q | A | E | * | R | R | G | V | F | I | I | * | G | R | C | I | | |
| 2 | C | P | S | * | I | A | * | R | G | F | H | H | L | R | T | <u>M</u> | <u>Y</u> | | |
| 1 | <u>M</u> | <u>P</u> | <u>K</u> | <u>L</u> | <u>N</u> | <u>S</u> | <u>V</u> | <u>E</u> | <u>G</u> | <u>F</u> | <u>S</u> | <u>S</u> | <u>F</u> | <u>E</u> | <u>D</u> | <u>D</u> | <u>V</u> | * | |
| 5' | ATGCCCAAGCTGAATAGCGTAGAGGGGTTTTTCATCATTTGAGGACGATGTATAA | | | | | | | | | | | | | | | | | 3' | |
| 3' | TACGGGTTCGACTTATCGCATCTCCCCAAAAGTAGTAAACTCCTGCTACATATT | | | | | | | | | | | | | | | | | 5' | |
| | H | G | L | Q | I | A | Y | L | P | K | * | * | K | L | V | I | Y | L | -1 |
| | | G | L | S | F | L | T | S | P | N | E | D | N | S | S | S | T | Y | -2 |
| | <u>A</u> | <u>W</u> | <u>A</u> | <u>S</u> | <u>Y</u> | <u>R</u> | <u>L</u> | <u>P</u> | <u>T</u> | <u>K</u> | <u>M</u> | <u>M</u> | Q | P | R | H | I | -3 | |

A reading frame from a start codon to the first stop codon is called an *open reading frame* - ORF (underlined above).